Quantifying Detection Probabilities for Proliferation Activities in Undeclared Facilities

Clemens Listner[1], Morton J. Canty[1], Gotthard Stein[2], Arnold Rezniczek[3] and Irmgard Niemeyer[1]

[1]Forschungszentrum Jülich, Germany
[2]Consultant, Bonn, Germany
[3]UBA GmbH, Herzogenrath, Germany

Abstract

International Safeguards is currently in an evolutionary process to increase effectiveness and efficiency of the verification system. This is an obvious consequence of the inability to detect the Iraq's clandestine nuclear weapons programme in the early 90's. By the adoption of the Programme 93+2, this has led to the development of Integrated Safeguards and the State-level concept. Moreover, the IAEA's focus was extended onto proliferation activities outside the State's declared facilities. The effectiveness of safeguards activities within declared facilities can and have been quantified with respect to costs and detection probabilities. In contrast, when verifying the absence of undeclared facilities this quantification has been avoided in the past because it has been considered to be impossible. However, when balancing the allocation of budget between the declared and the undeclared field, explicit reasoning is needed why safeguards effort is distributed in a given way. Such reasoning can be given by a holistic, information and risk-driven approach to Acquisition Path Analysis comprising declared and undeclared facilities [1]. Regarding the input, this approach relies on the quantification of several factors, i.e. costs of attractiveness values for specific proliferation activities, potential safeguards measures and detection probabilities for these measures also for the undeclared field. In order to overcome the lack of quantification for detection probabilities in undeclared facilities, the authors of this paper propose a general verification error model. Based on this model, four different approaches are explained and assessed with respect to their advantages and disadvantages: the analogy approach, the Bayes approach, the frequentist approach and the process approach. The paper concludes with a summary and an outlook on potential future research activities.

1 Introduction

Since the first ideas for supervising nuclear material, the verification system has evolved constantly. After having had first experiences with item-specific safeguards according the commitments in INFCIRC/66, the system of international safeguards was established by the signature and ratification of the Non-proliferation Treaty (NPT) in 1970. The treaty implementation has mainly been governed by comprehensive safeguards agreements (CSA) and

later the additional protocol (AP) with Integrated Safeguards. Until today this evolution of verifying treaty compliance has continued under a holistic approach called the State-level concept (SLC). The SLC's main idea is to go away from material centric approaches to a system analysis view of nuclear proliferation which clearly identifies the actors, their possibilities and their risks. Due to its general and comprehensive nature, the SLC has great potential to replace voluntary offer agreements (VOA) in nuclear weapon States (NWS) and to be used in other fields of treaty verification.
Underneath the new paradigmatic view to nuclear verification, the State-level concept essentially consists of three processes which help developing State-level safeguards approaches (SLA) [2]:

1. Identification of plausible acquisition paths.
2. Specification and prioritization of State-specific technical objectives (TO).
3. Identification of safeguards measures to address the technical objectives.

This paper concentrates on the first step which is also known as acquisition path analysis (APA). APA is defined as the analysis of all plausible sequences of activities which a State could consider to acquire weapons usable material [3]. The purpose of an APA is to determine whether a proposed set of safeguards measures is sufficient. Therefore, some overlap to the second step, the specification and prioritization of technical objectives, is obvious.
The approach to acquisition path analysis used in this paper has evolved over the past years [4, 1, 5]. Motivated from the fact, that the SLC tries to come up with adaptive safeguards approaches, the main idea of this approach to APA is to account for differentiation without discrimination. In order to accomplish this, the given information is processed in an objective, transparent, reproducible, standardized and well-documented way in contrast to classical reasoning-with-words- or black-box-approaches.
In order to fulfill these requirements, the methodology needs to quantify inter alia non-detection probabilities of proliferation activities. While this has been successfully accomplished in the case of nuclear material diversion in declared facilities [see 6], quantifying these probabilities for proliferation activities outside declared facilities is an unsolved task. Therefore, this paper will present four approaches to accomplish this.
In the following, the methodology and its recent enhancements will be presented. Then, the relation between the graph theoretic outcomes and the strategic assessment part will be explained in detail. Afterwards, four possibilities for quantifying non-detection probabilities will be presented. Finally, a conclusion and an outlook on future work will be given.

2 Materials and Methods

This paper's approach to acquisition path analysis consists of three general steps: First, the potential acquisition network is modeled based on the IAEA's physical model and experts' evaluations. Then, using this model all plausible acquisition paths are extracted automatically. Finally, the State's and the inspectorate's options are assessed strategically. The workflow is depicted in Figure 1. In the following, a description of the three stages will be given. A more in depth discussion can be found in Listner, Canty, Niemeyer, Rezniczek, and Stein [7].
During the first step of the process, also known as network modeling, a state-specific acquisition model is set up. Mathematically, such a network can be seen as a graph with material

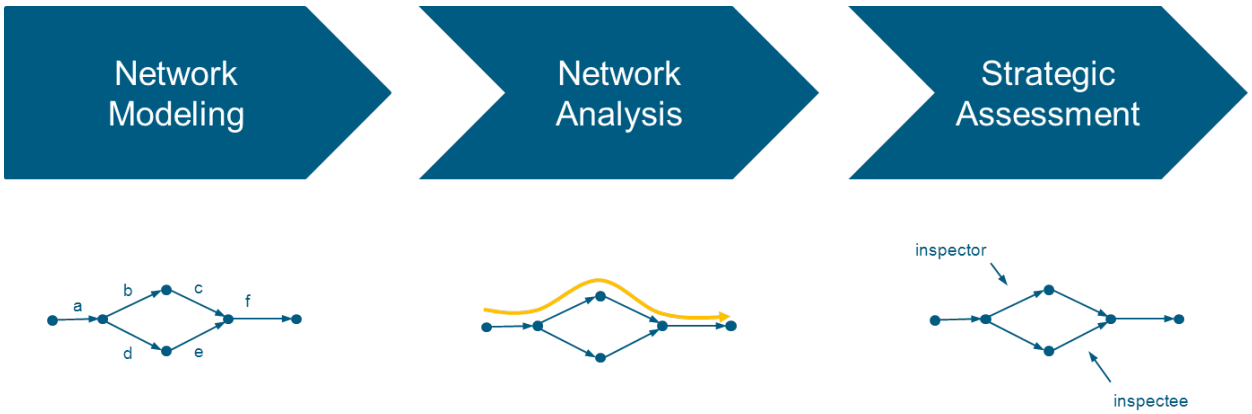

Figure 1. Three step approach to acuisition path analysis.

forms represented by nodes and processes represented by edges. As a starting point serves the IAEA's physical model [8] where all proliferation relevant materials and processes are formally described in a general acquisition model for nuclear weapons usable material.[1]
There are four categories of processes in the model: diversion from existing facilities (div), undeclared import (imp), misuse of existing facilities (mis), processing in clandestine facilities (cland). When assessing a State's options for acquiring nuclear weapons usable material, specific processes of these four types are put in or left out of the model. E.g. if a State does not have an enrichment facility on its ground, all edges of type misuse in connection with enrichment will be removed from the model. On the other hand, there will be always the option for enriching in clandestine facilities and hence these processes will remain in every State's case.
Besides the mere presence of edges in the model, these edges will be assessed in terms of attractiveness for the particular State. Therefore, three dimensions of attractiveness are used which originate in the GIF methodology [9]: Technical Difficulty (TD), Proliferation Time (PT) and Proliferation Cost (PC).[2] For each process and each dimension grades are given based on expert judgment. The grades range from 0 meaning a very attractive option to 3 being very unattractive. Using the arithmetic mean, for each edge e a single edge weight w_e is calculated from these figures.
After having specified the edge weights, it is necessary to model the inspectorate's side i.e. the possible technical objectives t with their respective non-detection probability $\beta_e^{(t)}$ on certain edge e. Also the inspectorate costs c_t generated by technical objective t have to be quantified. Although no specific safeguards measures have been determined at this point, an expert can estimate the costs for attaining a given detection probability based on experience and knowledge about a State's capabilities, fuel cycle as well as existing safeguards approaches. While these figures can be specified for the edges related to the declared fuel cycle, i.e. misuse and diversion, deriving this information for the undeclared, i.e. undeclared import and clandestine processing, is yet an unsolved task. Nevertheless, in this work it is assumed that such a quantification can be done in principle for all types of processes whether in declared facilities or elsewhere in the State.

[1] In principle, also the weaponization step itself could be modeled using a graph theoretic approach. However, due to the definition of acquisition path analysis given in International Atomic Energy Agency (IAEA) [3], this paper's approach ends at weapons usable material.

[2] These dimensions only represent technical aspects of proliferation as if no inspectorate was present. The interplay of State and inspectorate will be considered separately in the third stage of the process.

	No Alarm	**Alarm**
Compliant Behavior	$(0, 0)$	$(-f, -e)$
Non-compliant Behavior along path i	$(d_i, -c)$	$(-b, -a)$

Table 1. Game Theoretic Payoffs.

As a result of the first step, a directed multi-graph is produced that represents the State's options for producing weapons usable material including their attractiveness with respect to time cost and technical difficulty. Furthermore, also the inspectorate's options to control the activities are given including the costs and non-detection probabilities in specific areas of the State's acquisition network.
This graph is now analyzed in terms of all technically plausible acquisition paths. Therefore, a fully automated software using the depth first search algorithm extracts all paths from node 'Origin' to either node representing weapons usable material. For each path p_i, the overall attractiveness is calculated by the sum of the weights of the constituting edges $E(p_i)$, i.e.

$$l_i = \sum_{e \in E(p_i)} w_e. \tag{1}$$

The list of paths is then reordered by attractiveness and all paths are visualized. It has to be emphasized that not only the shortest path but all technically plausible paths. Therefore, this approach is comprehensive and avoids to ignore technically less attractive paths which could be strategically interesting.
Using the results of the first and second step, especially the list of paths with their respective attractiveness as well as the non-detection probabilities of technical objectives, the third steps strategically assesses the options for both parties, i.e. the State and the inspectorate. In order to accomplish this, all acquisition paths and the option of compliant behavior are considered to be the State's strategy set. On the other hand, the strategies of the IAEA are all combinations of technical objectives (TOC) that have been defined in the first part of the process. The overall non-detection probability of TOC_j for a given path p_i can be calculated using the product rule for probabilities by

$$\beta_{ij} = \prod_{e \in E(p_i), t \in TOC_j} \beta_e^{(t)}. \tag{2}$$

For each strategy combination a pair of payoff values for State and Inspectorate (H_1, H_2) can be defined (see Table 1). For the IAEA, the strategic outcomes in increasing order of preference are undetected non-compliance $(-c)$, detected non-compliance $(-a)$, false alarm $(-e)$ and compliance without alarm (0). These parameter can be selected freely as long as the ordering is kept.
Regarding the State, the strategic outcomes ordered increasingly by preference are detected non-compliance $(-b)$, false alarm $(-f)$, compliance without alarm (0) and successful acquisition along path i (d_i). The path length l_i calculated in step two is used to obtain the payoff values for successful acquisition by

$$d_i = \frac{l_1}{l_i}. \tag{3}$$

The decision whether an alarm is raised by the inspectorate depends probabilistically on the non-detection probabilities. Hence, for each strategy combination an expected outcome

for both players can be calculated. In case the State decides to follow an acquisition path i and the IAEA has TOC j in place, this payoff for the State is given by the expected benefit from a successful acquisition plus the risk of getting caught red-handed, i.e.

$$H_1^{(i)} = d_i\beta_{ij} - b(1-\beta_{ij}). \tag{4}$$

For the IAEA, the expected payoff can be derived from the sum of the risks of detected and undetected non-compliance , i.e.

$$H_2^{(i)} = -c\beta_{ij} - a(1-\beta_{ij}). \tag{5}$$

In case the State behaves in compliance with its given commitments, the outcome for both sides is only determined by the false alarm risk with false alarm probability α, i.e.

$$H_1^{(compliant)} = -f\alpha \tag{6}$$

for the State and

$$H_2^{(compliant)} = -e\alpha \tag{7}$$

for the IAEA.
Based on these considerations, a stable strategy combination (H_1^*, H_2^*) known as the Nash equilibrium can be calculated using the Lemke-Howson-algorithm [10]. The Nash equilibrium is characterized by the fact that its impossible for either of the two actors to deviate unilaterally from the equilibrium strategy and increase its expected payoff. Hence, it seems rational for both players not to deviate and pursue the equilibrium strategy. This very limited definition of rationality only means that the actors care for the risks and benefits in they are facing.
Using the equilibrium payoff value for the IAEA and scaling the IAEA's payoff parameters to $c = 1$, it is possible to define effectiveness as

$$E = 100\% + H_2^*. \tag{8}$$

In case of 0% effectiveness, the equilibrium ends in non-compliance with no possibility of detection. For 100% effectiveness, compliance with no false alarm is achieved almost surely. As the ultimate goal of acquisition path analysis is the selection of a TOC inducing compliant behavior (expressed by the term sufficient in the APA definition), this paper proposes to use a TOC leading to a high effectiveness value in the Nash equilibrium.
Moreover, in cases where compliant behavior can be induced in the Nash equilibrium, it is also possible and reasonable to gain an increase in efficiency. Iterating over a cost threshold W and calculating the Nash equilibrium for this range of values, gives strategy for which a given level of effectiveness can be achieved at minimum cost.

3 Four Approaches to Quantifying Detection Probabilities

The previous section has shown how the model can be used to determine an optimal set of technical objectives. As input on behalf of the State, the model requires an assessment of each process' attractiveness w_e and the resulting payoff values for each path d_i. On behalf of the inspectorate, for each technical objective t a cost estimate for each technical objective is

	No Alarm	Alarm
Compliance	1 – α	α
Non-Compliance	β	1 – β

Figure 2. Verification Error Matrix.

needed as well as an estimate of the non-detection probability $\beta_e^{(t)}$ for each process e given that a technical objective t is in place. It turns out that the attractiveness values, the cost estimates and the declared facilities' non-detection probabilities can be obtained relatively easily because there are models available for the estimation of these parameters. However, until now there are no such models for the estimation of non-detection probabilities for processes in covert facilities as well as undeclared import.
In the past, the estimation of such non-detection probabilities has been considered to be an impossible task. The reasons for the reluctance to quantify these parameters can be found in the lack of system boundaries of clandestine nuclear facilities as they can be located anywhere in a State. The same applies to the case of undeclared import, where the location of possible indicators could even be found worldwide. Moreover, it is not even clear which indicators could give the relevant hint to a clandestine facility.
All these problems seem to be good reasons to think about the detection of clandestine facilities and undeclared import only in a qualitative way. However, this would lead to the problem of how to justify the spending of budget on the detection of clandestine facilities against conventional safeguards measures whose effectiveness can be quantified very elegantly. A model calculating quantitative estimates for the non-detection probabilities can overcome this issue. Also, this problem is similar to effectiveness quantification in the intelligence realm and there has been research on how to address this [see 11].
In the past it has been shown that hypothesis testing is a powerful tool that can be applied in the context of treaty verification to estimate the errors [see 6]. It assumes that a State can either behave compliantly or not. On the other hand, the inspectorate has the possibility to raise an alarm or not. Thus, four event combinations result from this error model which are displayed in Figure 2. The main diagonal entries of this matrix stand for a properly working verification system which either raises an alarm if appropriate or does not if it would be inappropriate. The off-diagonal elements however reflect errors in the verification system. An error of the first kind, also known as a false alarm, will occur, if the State behaves compliantly but the inspectorate raises an alarm despite that fact. This error's probability is denoted by α. The error of the second kind is also known as non-detection of incompliance. This error will occur, if the State proliferates but the inspectorate is not able to detect this behavior and thus will not raise an alarm. This error's probability is denoted by β.
Based on this error model, the existing literature and developing new ideas, four possibilities will be presented how to estimate the non-detection probabilities in case of undeclared facilities or import. These suggestions should be seen as a starting point for further discussion and research.

3.1 Possibility A: The Analogy Approach

The first and by far the simplest possibility starts by looking into declared facilities. There, the safeguards system can obtain a non-detection probability of $\beta_{declared} = 10\%$ if all measures, like e.g. PIVs and IIVs, are in place. By analogy, the same non-detection probability of $\beta_{undeclared} = 10\%$ is assumed for undeclared facilities in case all measures, like e.g. open source information analysis taskings, are applied here as well. If only parts of the measures are applied, a linear scaling procedure increases the non-detection probability. E.g. in case only half of the measures are applied, the detection probability reduces from $1 - \beta_{undeclared} = 90\%$ to $1 - \beta_{undeclared} = 45\%$.
This approach gives a model which is very simple and easy to understand. However, a validation of the stated non-detection probabilities is merely impossible.

3.2 Possibility B: The Bayesian Approach

The second approach uses Bayes' theorem to model the information analysis process and then estimates the detection probability from a simulation step. In this context, the event A_j means that the proliferation activity j, e.g. the use of a clandestine reprocessing facility, is carried out by the State. $B = \{B_1, \ldots, B_n\}$ represent the set of available information pieces. Based on these probabilistic events, the Bayes formula retrieves the probability of a proliferation activity A_j given a set of available information B as

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{P(B|A_j)P(A_j) + P(B|\overline{A}_j)P(\overline{A}_j)} \tag{9}$$

In this formula, the probabilities $P(B|A_j)$ can be derived from the physical model which lists indicators, i.e. pieces of information, with their probability of occurrence in case a specific proliferation activity is carried out. The probabilities given the complementary events $P(B|\overline{A}_j)$ would have to be estimated by experts in a similar way. An open issue remains how the prior probabilities $P(A_j)$ and $P(\overline{A}_j)$ could be obtained.
Once the Bayes formula is applied to derive the probability $P(A_j|B)$, the information analysis process would raise an alarm, if this probability exceeds a given threshold T. In order to derive the non-detection probabilities β, one checks the correctness of the information analysis process for any combination of B, A_j and $\overline{A}_j$ weighted by the probability of each event combination. The error of the first second kind then gives the non-detection probability β. Again, estimating the prior probability of each event combination remains an unsolved problem.
As a conclusion, one can say that the Bayesian approach helps structuring the problem of quantifying detection probabilities in a qualitative environment. Moreover, the physical model already includes certain information which can serve as input. However, it is a non-trivial task to obtain the prior probability of a proliferation activity. In order to be non-discriminatory, the methodology would have to assume the same priors for each State although this hardly reflects reality.

3.3 Possibility C: The Frequentist Approach

As a third possibility, historical events in the field of non-proliferation can be used to retrieve estimates for the non-detection probability. Therefore, the error matrix is filled with the absolute number of events (see Figure 3). Using these figures, the non-detection probability

	No Alarm	Alarm
Compliance	$H_{compliance\ without\ alarm}$	$H_{false\ alarm}$
Non-Compliance	$H_{undetected\ noncompliance}$	$H_{successful\ detection}$

Figure 3. Estimated Verification Error Matrix.

can be estimated using

$$\hat{\beta} = \frac{H_{\text{undetected non-compliance}}}{H_{\text{undetected non-compliance}} + H_{\text{succesful detection}}}. \tag{10}$$

Similarly, an estimate for the false alarm probability can be given by

$$\hat{\alpha} = \frac{H_{\text{false alarm}}}{H_{\text{false alarm}} + H_{\text{compliance without alarm}}}. \tag{11}$$

In practice, the number of events can be obtained from the safeguards implementation report or other sources of information. Also, one could think of aggregating the data using different criteria such as counting only events that took place in a single year, that refer to a particular State or that cover a specific proliferation activity.
The advantages of this approach result from the strong quantitative basis and the simplicity because only counting events is required. However, the disadvantage of relatively few data points for non-compliance are obvious. This could be a source of error.

3.4 Possibility D: The Process Approach

Finally, the fourth approach considers α and β to be "'measurement errors"' of the inspectorate's information analysis process. This information analysis process can be subdivided into five components aüccording to the intelligence cycle [see 12]: plan, collect, process, analyze, disseminate.
For each subprocess j, this approach estimates the errors for a false alarm, α_j, and non-detection, β_j, based on the error sources within the respective subprocesses. Assuming independence of error probabilities among the subprocesses, the overall errors can then be calculated as

$$\alpha_{total} = 1 - \prod_{j=1}^{5} 1 - \alpha_j \tag{12}$$

and

$$\beta_{total} = 1 - \prod_{j=1}^{5} 1 - \beta_j. \tag{13}$$

An advantage of this approach is the fact that it helps structuring the problem of estimating verification error probabilities despite the absence of complete error models. It also gives hints where to improve the information analysis process. However, the quantification of errors is still necessary on a lower level. This is not easy to accomplish for all subprocesses of the information analysis process.

4 Conclusions and Outlook

This paper has shown how acquisition path analysis can be carried out using a comprehensive methodology which is yet compatible with the principles defined in Cooley [2]. Furthermore, a possibility for determining technical objectives has been proposed. This approach delivers a set of technical objectives with optimal effectiveness under the assumptions of a game theoretic model. Besides the high degree of automation, this approach also allows for an inherent randomization of technical objectives. However, the analyst has to specify a set of parameters in this approach. Therefore a good understanding of the model is necessary and the influence of the parameters on the model's outcome is very complex.
A major point of criticism of this methodology has been the question how to quantify the non-detection probabilities of proliferation activities outside declared facilities. As a starting point for discussion, this paper has outlined four approaches how this quantification could be implemented.
In the future, further case studies will have to be carried out with respect to the continuous improvements of the methodology. This will include the analysis of the outcome's sensitivity on the selected parameters. In these case studies, special focus will be given to the question how other areas, like the verification of disarmament treaties, can benefit from this approach. Also, further work will be carried out regarding the quantification of non-detection probabilities. Finally, the methodology will be iteratively improved with the help of experts at the IAEA.

5 Acknowledgments

This paper was prepared as an account of work sponsored by the Government of the Federal Republic of Germany within the Joint Programme on the Technical Development and Further Improvement of IAEA Safeguards between the Federal Republic of Germany and the IAEA.

References

[1] C. Listner, M. J. Canty, A. Rezniczek, G. Stein, and I. Niemeyer. "Approaching acquisition path analysis formally - experiences so far". In: *Proceedings of the 54th INMM Annual Meeting*. INMM. 2013.

[2] J. N. Cooley. "Progress in Evolving the State-level Concept". In: *Seventh INMM/ESARDA Joint Workshop Future Directions for Nuclear Safeguards and Verification*. 2011.

[3] International Atomic Energy Agency (IAEA). "IAEA Safeguards Glossary". In: *International Nuclear Verification Series No. 3* (2001).

[4] C. Listner, M. J. Canty, A. Rezniczek, G. Stein, and I. Niemeyer. "A Concept for Handling Acquisition Path Analysis in the Framework of IAEA's State-level Approach". In: *Proceedings of the 53rd INMM Annual Meeting*. INMM. 2012.

[5] C. Listner, M. J. Canty, A. Rezniczek, G. Stein, and I. Niemeyer. "Approaching acquisition path analysis formally - a comparison between AP and non-AP States". In: *Proceedings of the 35rd ESARDA Annual Meeting*. 2013.

[6] Rudolf Avenhaus and Morton John Canty. *Compliance quantified*. Cambridge University Press Cambridge, 1996.

[7] C. Listner, M. J. Canty, I. Niemeyer, A. Rezniczek, and G. Stein. *A Concept for Handling Acquisition Path Analysis in the Framework of IAEA's State-level Approach*. Tech. rep. JOPAG/04.13-PRG-400. 2013.

[8] International Atomic Energy Agency (IAEA). "The Physical Model". STR-314. 1999.

[9] GEN IV International Forum. *Evaluation Methodology for Proliferation Resistance and Physical Protection of Generation IV Nuclear Energy Systems*. 2006.

[10] M. J. Canty. *Resolving conflicts with Mathematica: algorithms for two-person games*. AP, 2003.

[11] P. Lehner, A. Michelson, and L. Adelman. *Measuring the Forecast Accuracy of Intelligence Products*. Tech. rep. MITRE, 2010.

[12] *The Intelligence Cycle*. Retrieved from `https://www.cia.gov/kids-page/6-12th-grade/who-we-are-what-we-do/the-intelligence-cycle.html`. Central Intelligence Agency. Mar. 2013.

Nomenclature

AP	additional protocol
APA	acquisition path analysis
cland	processing in clandestine facilities
CSA	comprehensive safeguards agreements
div	diversion from existing facilities
imp	undeclared import
mis	misuse of existing facilities
NPT	Non-proliferation Treaty
NWS	nuclear weapon State
PC	Proliferation Cost
PT	Proliferation Time
SLA	State-level approach
SLC	State-level concept
TD	Technical Difficulty
TO	technical objectives
TOC	technical objectives combination
VOA	voluntary offer agreement